Modeling Functional Responses with Multivariate Covariates

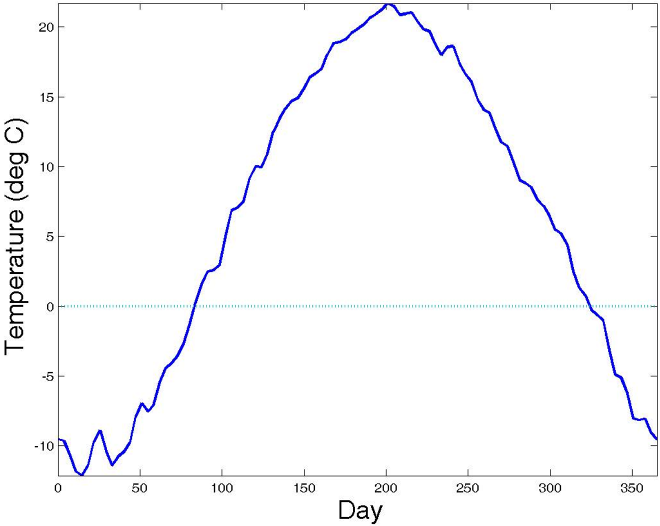
Pang Du

Department of Statistics
Virginia Tech

The average Canadian weather data

- We have 35 weather stations distributed across four climate zones:
    - Atlantic (16)
    - Pacific (6)
    - Continental (13)
    - Arctic (4)
- The dependent variable is Temp(t), a function representing daily temperatures averaged over 1960-1994.
- The temperature functions were obtained by expanding the original 365 discrete daily averages in terms of 65 Fourier basis functions.

Montreal's temperature profile



The functional ANOVA model

The model for the *m*th temperature function in the *g*th zone is

$$\text{Temp}_{mg}(t) = \mu(t) + \alpha_g(t) + \epsilon_{mg}(t).$$

- $\mu(t)$: the grand mean function across all 35 weather stations.
- $\alpha_g(t)$: the specific effects on temperature of being in climate zone g.
- $\epsilon_{mg}(t)$: the residual function showing unexplained variation specific to the *m*th weather station within climate group g.
- To be able to identify $\alpha_g(t)$ uniquely, we require that they satisfy the zero sum constraint:

$$\sum \alpha_g(t) = 0 \text{ for all } t.$$

Model setup

- Design matrix $\mathbf{Z}$: a 35 by 5 matrix with column 1 containing all 1's, and columns $g+1, g=1,...,4$ containing the group indicate vector in each row.
- **Append a final row with 0 in column 1, and 1's in the remaining columns.**
- Functional response vector $\mathbf{Temp}(t)$: contain the 35 temperature profiles **plus a final function that is zero for all** t.
- Functional regression coefficient vector $\boldsymbol{\beta}(t)$: contain the functions $(\mu(t), \alpha_1(t), ..., \alpha_4(t))$.
- The model in matrix notation is

$$\mathbf{Temp}(t) = \mathbf{Z}\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t),$$

Fitting the model

- The residual $\text{Temp}_i(t) - \mathbf{Z}_i \boldsymbol{\beta}(t)$ is now a function.
- The least squares fitting criterion becomes

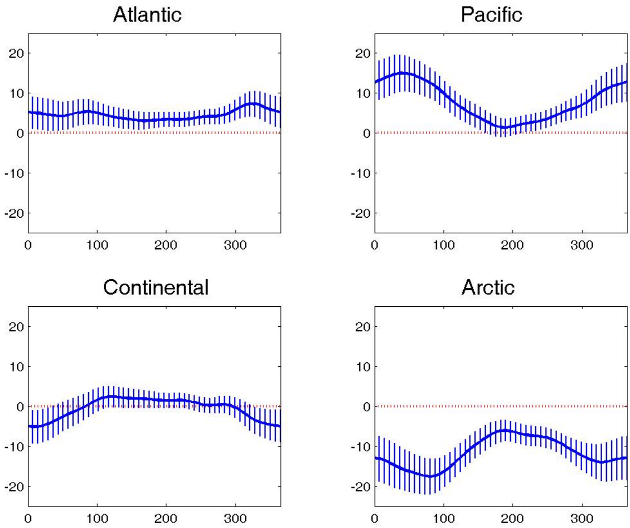
$$\text{LMSSE}(\boldsymbol{\beta}) = \sum_{g=1}^{4} \sum_{m=1}^{N_g} \int [\text{Temp}_{mg}(t) - \sum_{j=1}^{g+1} z_{(mg),j} \beta_j(t)]^2 dt.$$

subject to the constraint $\sum_{j=2}^{5} \beta_j(t) = 0$.

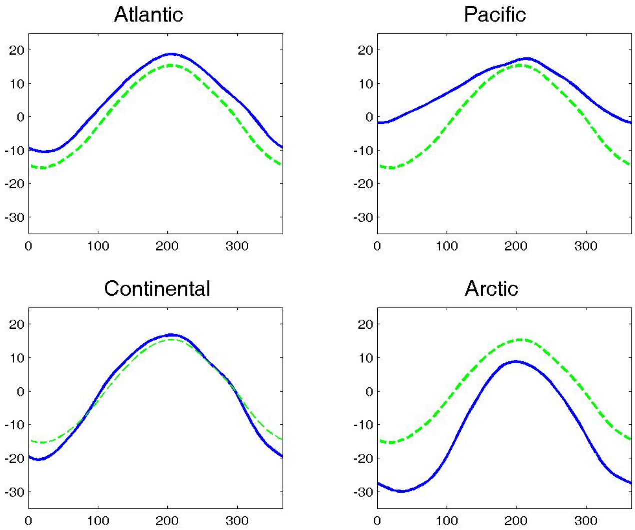
- This is minimized with respect to the regression functions by

$$\hat{\boldsymbol{\beta}}(t) = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \mathbf{Temp}(t).$$

The region effects $\alpha_g(t)$



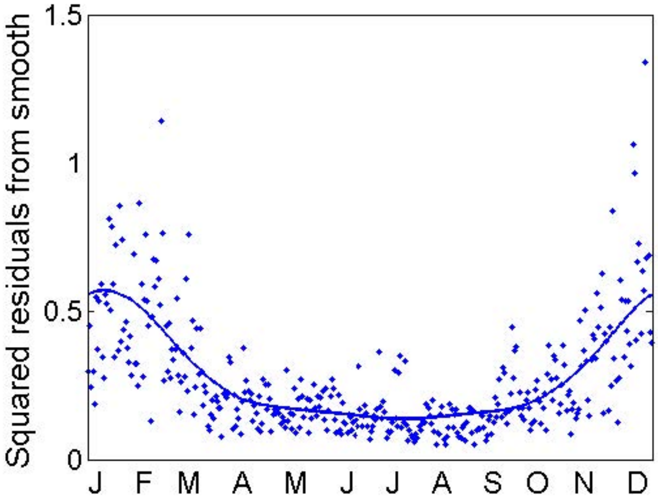
The region means $\mu(t) + \alpha_g(t)$



What do we see?

- Atlantic: a consistent 5°C warmer than the Canadian average.
- Pacific: summer temperature close to the Canadian average but much warmer winter temperatures.
- Continental: slightly warmer summer temperatures than the average but colder in the winter by about 5°C.
- Arctic: colder than the Canadian average, even more so in March than in January.

The estimated standard error function $\hat{\sigma}(t)$



Assessing the fit: Motivation

- Is there significant variation in temperature over climate zones? Of course there is! This does not seem like an interesting question.
- On the other hand, whether the Atlantic, Pacific and Continental stations are significantly different in the summer might be.
- Interesting summaries of fit, of effects, and inferences are likely to be **local** in nature.

Assessing the fit: Sums and means of squares

- It is useful to use the error sum of squares function
$$\mathrm{SSE}(t) = \sum_{m,g} [\mathrm{Temp}_{mg}(t) - \mathbf{Z}_{mg}\hat{\beta}(t)]^2.$$
to assess fit at or near time t.
- As in ordinary regression, we can compare this to the variation of the response about its mean
$$\mathrm{SSY}(t) = \sum_{m,g} [\mathrm{Temp}_{mg}(t) - \hat{\mu}(t)]^2.$$
- The mean squared error functions:
$$\mathrm{MSE}(t) = \mathrm{SSE}(t)/\mathrm{df}(\mathrm{error})$$
$$\mathrm{MSR}(t) = \frac{\mathrm{SSY}(t) - \mathrm{SSE}(t)}{\mathrm{df}(\mathrm{model})}$$

Multiple correlation and F-ratio functions

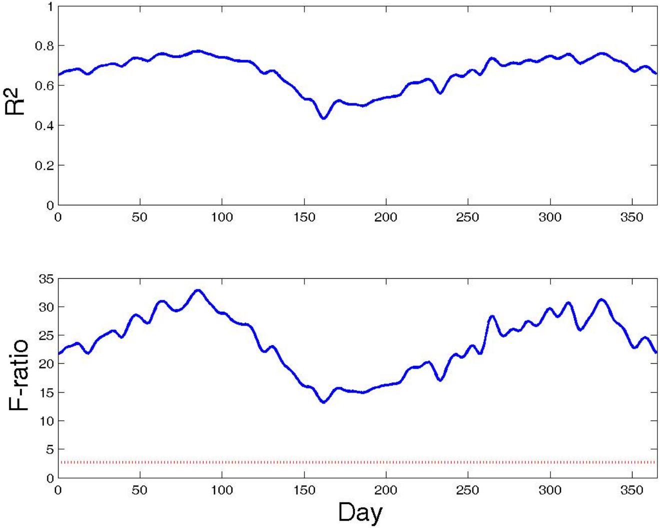
- The squared multiple correlation function is

$$\mathrm{RSQ}(t) = [\mathrm{SSY}(t) - \mathrm{SSE}(t)]/\mathrm{SSY}(t).$$

- and the F-ratio function is

$$\mathrm{FRATIO}(t) = \frac{\mathrm{MSR}(t)}{\mathrm{MSE}(t)}.$$

R^2 and F-ratio plots



What do we see?

- Relatively high R^2.
- F-ratio: everywhere substantially above the 5% significance level of 2.92.
- Differences between the climate zones are substantially stronger in the spring and autumn, rather than in the summer and winter as we might expect.

Goals

- We want a general framework for estimating functional parameters in this and other linear models.
- We want to be able to penalize the roughness of any functional parameter β_j.
- We also want the capacity to estimate confidence intervals for a parameter,
- and for functionals $\rho(\beta_j)$ of a parameter.

Basis expansion for $\beta_j(t)$

- Let the regression coefficient vector $\boldsymbol{\beta}(t)$ have the expansion
$$\boldsymbol{\beta}(t) = \mathbf{B}\boldsymbol{\theta}(t),$$
where matrix $\mathbf{B}$ is q by K_β and the K_β basis functions $\theta_l(t)$ are contained in vector $\boldsymbol{\theta}(t)$.
- In the temperature example, it would be natural to use a certain number K_β of Fourier basis functions.

Roughness penalty on $\beta_j(t)$

- If the response curves in $\mathbf{y}(t)$ are rough, we may want to impose some smoothness on the estimated β_j's.
- Let L be a linear differential operator, such as $L = D^2$, that defines the variation $L\boldsymbol{\beta}(t)$ we wish to penalize.
- Our roughness penalty on $\boldsymbol{\beta}(t)$ is

$$\mathrm{PEN}(\boldsymbol{\beta}) = \int [L\boldsymbol{\beta}(s)]'[L\boldsymbol{\beta}(s)]ds.$$

The penalized least squares criterion

- Let the response function vector $\mathbf{y}(t)$ have the basis function expansion in terms of K_y basis functions $\phi_k(t)$:

$$\mathbf{y}(t) = \mathbf{C}\boldsymbol{\phi}(t).$$

- Note that the basis functions for $\mathbf{y}$ and $\boldsymbol{\beta}$ may differ.
- Then the penalized least squares function is

$$\text{PENSSE}(\mathbf{y}|\boldsymbol{\beta}) = \int (\mathbf{C}\boldsymbol{\phi} - \mathbf{Z}\mathbf{B}\boldsymbol{\theta})'(\mathbf{C}\boldsymbol{\phi} - \mathbf{Z}\mathbf{B}\boldsymbol{\theta}) + \lambda \int (L\mathbf{B}\boldsymbol{\theta})'(L\mathbf{B}\boldsymbol{\theta}).$$

- Can incorporate a weight matrix $\mathbf{W}$ if needed.

PLS in matrix form

- Define the matrices:

$$\mathbf{J}_{\phi\phi} = \int \boldsymbol{\phi}\boldsymbol{\phi}', \mathbf{J}_{\theta\theta} = \int \boldsymbol{\theta}\boldsymbol{\theta}', \mathbf{J}_{\phi\theta} = \int \boldsymbol{\phi}\boldsymbol{\theta}'$$

- and the roughness penalty matrix

$$R = \int (L\boldsymbol{\theta})(L\boldsymbol{\theta})'.$$

- The fitting criterion now can be expressed as

$$\begin{aligned}\mathrm{PENSSE}(\mathbf{y}|\beta) = \mathrm{trace}(\mathbf{C}'\mathbf{C}\mathbf{J}_{\phi\phi}) + \mathrm{trace}(\mathbf{Z}'\mathbf{Z}\mathbf{B}\mathbf{J}_{\theta\theta}\mathbf{B}') \\ - 2\mathrm{trace}(\mathbf{B}\mathbf{J}_{\theta\phi}\mathbf{C}'\mathbf{Z}) + \lambda\mathrm{trace}(\mathbf{B}\mathbf{R}\mathbf{B}').\end{aligned}$$

Kronecker product and vec notation

- The **Kronecker product** $\mathbf{A} \otimes \mathbf{B}$ of two matrices $\mathbf{A}$ and $\mathbf{B}$ is the matrix generated from $\mathbf{A}$ with each entry a_{ij} replaced by a submatrix $a_{ij}\mathbf{B}$.
- Let $\text{vec}(\mathbf{A})$ be the mn-vector generated from a matrix $\mathbf{A}_{m\times n}$ by writing $\mathbf{A}$ as a vector column-wise.
- A useful identity:

$$\text{vec}(\mathbf{ABC}') = (\mathbf{C} \otimes \mathbf{A})\text{vec}(\mathbf{B}),$$

The normal equation for B

- Taking the matrix derivative with respect to $\mathbf{B}$ and setting it to 0 gives

$$(\mathbf{Z}'\mathbf{Z}\mathbf{B}\mathbf{J}_{\theta\theta} + \lambda\mathbf{B}\mathbf{R}) = \mathbf{Z}'\mathbf{C}\mathbf{J}_{\phi\theta}.$$

- We can use the Kronecker product and the vec notation to convert the normal equation to

$$[\mathbf{J}_{\theta\theta} \otimes (\mathbf{Z}'\mathbf{Z}) + \mathbf{R} \otimes \lambda\mathbf{I}]\text{vec}(\mathbf{B}) = \text{vec}(\mathbf{Z}'\mathbf{C}\mathbf{J}_{\phi\theta}).$$

- The estimate $\hat{\mathbf{B}}$ is therefore

$$\begin{aligned}\text{vec}(\hat{\mathbf{B}}) &= [\mathbf{J}_{\theta\theta} \otimes (\mathbf{Z}'\mathbf{Z}) + \mathbf{R} \otimes \lambda\mathbf{I}]^{-1}\text{vec}(\mathbf{Z}'\mathbf{C}\mathbf{J}_{\phi\theta}) \\ &= \mathbf{S}_\beta\text{vec}(\mathbf{C}).\end{aligned}$$

Some variations

- A separate smoothing parameter λ_j for each regression function $\beta_j(t)$: replacing $\lambda\mathbf{I}$ by $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_q)$ gives

$$\text{vec}(\hat{\mathbf{B}}) = [\mathbf{J}_{\theta\theta} \otimes (\mathbf{Z}'\mathbf{Z}) + \mathbf{R} \otimes \Lambda]^{-1}\text{vec}(\mathbf{Z}'\mathbf{C}\mathbf{J}_{\phi\theta}).$$

- Constrain predicted response function $\hat{\mathbf{y}}$ to be smooth rather than $\boldsymbol{\beta}$: replacing $\mathbf{I}$ by $\mathbf{Z}'\mathbf{Z}$ gives

$$\text{vec}(\hat{\mathbf{B}}) = [(\mathbf{J}_{\theta\theta} + \lambda\mathbf{R}) \otimes (\mathbf{Z}'\mathbf{Z})]^{-1}\text{vec}(\mathbf{Z}'\mathbf{C}\mathbf{J}_{\phi\theta}).$$

Fitting the raw data: the PLS

- Suppose that the response variable $\mathbf{y}(t)$ is not pre-smoothed but in the raw data matrix form $\mathbf{Y}$.
- $\mathbf{Y}$ is an N by m matrix whose ith row equals $y_i(t_1), \ldots, y_i(t_m))$ with $t_1, \ldots, t_m$ being the sampling points for the response functions.
- Let Θ be the m by K_β matrix whose jth row is $\boldsymbol{\theta}(t_j)$ with $\boldsymbol{\theta}$ being the basis functions for $\boldsymbol{\beta}$.
- The penalized least squares criterion becomes

$$\|\mathbf{Y} - \mathbf{ZB}\Theta'\|^2 + \lambda \|\mathbf{L}\boldsymbol{\beta}(t)\|^2.$$

Fitting the raw data: the estimate

- The normal equation is

$$(\mathbf{Z}'\mathbf{Z})\mathbf{B}(\Theta'\Theta) + \lambda\mathbf{B}\mathbf{R} = \mathbf{Z}'\mathbf{Y}\Theta.$$

- The estimate of **B** is

$$\mathrm{vec}(\hat{\mathbf{B}}) = [(\Theta'\Theta) \otimes (\mathbf{Z}'\mathbf{Z}) + \mathbf{R} \otimes \lambda\mathbf{I}]^{-1}\mathrm{vec}(\mathbf{Z}'\mathbf{Y}\Theta').$$

Motivation

- Estimating the entire regression function $\beta_j(t)$ is fine,
- but we want to focus our attention on local or specific shape features of $\beta_j(t)$.
- Perhaps, for example, we want to examine the behavior of the temperature coefficient functions in mid-winter.

Definition

- A *functional probe* or *functional contrast* is of the form
$$\rho(\beta_j) = \int \xi(s)\beta_j(s)ds.$$
- $\xi(s)$ is a weight function that we choose so as to concentrate our attention on a local region, or to look for specific patterns of variation in $\beta_j(t)$.
- There is no particular need for $\xi(s)$ to integrate to 0.
- When $\beta_j(s)$ has the basis function expansion $\beta_j(s) = \mathbf{B}_j\boldsymbol{\theta}(s)$, where $\mathbf{B}_j$ is the jth row of $\mathbf{B}$, the contrast becomes
$$\rho(\beta_j) = \mathbf{B}_j \int \xi(s)\boldsymbol{\theta}(s)ds.$$

Some examples

Fixed a time point t,

- **Point evaluation**:

$$\xi(s) = \delta(s - t),$$

where $\delta(s) = \infty$ at $s = 0$ and 0 elsewhere is the Dirac's delta function.
This simply produces the function value $\beta_j(t)$.

- **Local behavior**: we can use

$$\xi(s) = \exp[(s - t)^2/(2\sigma)]$$

to assess the behavior of β_j in a neighborhood of t of a size determined by the constant σ.

How do I work out CLs for these probes?

- The random element in a linear model is the smoothing residual function value

$$\epsilon_{ij} = y_{ij} - x_i(t_j).$$

- Any linear function of the data (including the smoothed data) inherits its variance from the variance of the data.
- The variance of the data conditional on the smoothing model is the variance of the residuals.

Two tasks

- Estimate the variance of the *smoothing residuals* for a single response (the mean can usually be taken to be 0.) Let's call this estimate Σ_e.
- Assuming independence of the observations, the variance of the whole response data matrix is

$$\mathrm{Var}[\mathrm{vec}(\mathbf{Y})] = \Sigma_e \otimes \mathbf{I}.$$

- Work out the linear mapping from the data to the probe $\rho(\beta_j)$ that is being estimated. Let us call this $\mathbf{M}_j$.

The rest is easy:

$$\mathrm{Var}[\rho(\beta_j)] = \mathbf{M}_j'(\Sigma_e \otimes \mathbf{I})\mathbf{M}_j.$$

How do I work out mapping $\mathbf{M}_j$?

In the examples given, $\rho(\beta_j)$ is three linear mappings moving from the data:

- The linear mapping from the raw data matrix $\mathbf{Y}$ to the coefficient matrix $\mathbf{C}$ defining the smooth functions in $\mathbf{y}(t)$. Using the "hat" matrix $\mathbf{S}_y$, this is

$$\text{vec}(\mathbf{C}) = (\mathbf{S}_y \otimes \mathbf{I})\text{vec}(\mathbf{Y}).$$

- The linear mapping from $\mathbf{C}$ to the regression coefficient function coefficient vector $\mathbf{B}_j'$. We worked this out already, and called it $\mathbf{S}_\beta$.
- The linear mapping from $\mathbf{B}_j'$ to the value of the probe. This is

$$\mathbf{U} = \int \xi(s)\boldsymbol{\theta}'(s)ds.$$

The mapping $\mathbf{M}_j$

- Now we have it, namely

$$\mathbf{M}_j = \mathbf{U}_j \mathbf{S}_\beta (\mathbf{S}_y \otimes \mathbf{I}).$$

- This process is easy to extend to probes $\xi(s)$ involving all regression coefficients.
- For example, the variance of vec$[\hat{\beta}(\mathbf{t})]$ where $\mathbf{t}$ is a vector of values of t, is

$$(\Theta \otimes \mathbf{I})\mathbf{S}_\beta (\mathbf{S}_y \otimes \mathbf{I})(\Sigma_e \otimes \mathbf{I})(\mathbf{S}_y' \otimes \mathbf{I})\mathbf{S}_\beta' (\Theta \otimes \mathbf{I})'.$$

where Θ is the matrix of values of $\boldsymbol{\theta}$ at $\mathbf{t}$.

Some cautionary notes

- These sampling variances would only be "exact" if we knew Σ_e. The value of our confidence limit estimates depends critically on the quality of the estimate of Σ_e (many open questions!).
- We are assuming that the distribution of a probe is well summarized by its mean and variance.
- Our estimates are all conditioned on how many basis functions we use for both $y_i(t)$ and $\beta_j(t)$, namely K_y and K_β. Since we never know exactly how many to use, these should be regarded as random quantities, and a Bayesian treatment seems to be indicated.
- We should back up the use of these "delta method" confidence regions by bootstrapping and simulations.

- Regressing a functional response on multivariate independent variables or on a design matrix is not much different from the conventional regression analysis.
- One important difference is that we want to do local inference and interval estimation.
- We have, too, the capacity to smooth estimated functional parameters.
- But the number of basis functions that we use is not a fixed parameter in the traditional sense.